5-2015

Investigation of the thermodynamic properties of the oligomerization domain of heterogeneous nuclear ribonucleoprotein c

John P. O'Brien III
University of Tennessee at Chattanooga, jyd945@mocs.utc.edu

Investigation of the Thermodynamic Properties of the Oligomerization Domain of Heterogeneous Nuclear Ribonucleoprotein C

John Patrick O'Brien III

Departmental Honors Thesis
University of Tennessee at Chattanooga
Department of Physics, Geology, and Astronomy

Project Director: Dr. Kristin B. Whitson
Examination Date: Wednesday, April 1st, 2015

Members of the Examining Committee:

Dr. Tatiana Allen
Dr. Bob Marlowe
Dr. Kristin B. Whitson
Dr. Pam Carter
Dr. Stefanie Whitson

Dr. Kristin B. Whitson, Project Director

Dr.Bob Marlowe, Department Examiner

Dr. Tatiana Allen, Department Examiner

Dr. Pam Carter, Liaison, Departmental Honors Committee

Dr. Stefanie R. Whitson, Interdisciplinary Examiner, Chemistry Department

---

Chairperson, Departmental Honors Committee

"It is by logic that we prove, but by intuition that we discover"

Henri Poincaré, French Mathematician, 19th century

Abstract

Investigation of the Thermodynamic Properties of the Oligomerization Domain of Heterogeneous Nuclear Ribonucleoprotein C

by John Patrick O'Brien III

The heterogeneous nuclear ribonucleoprotein C (hnRNP C) performs a critical role in the processing of nascent pre-messenger ribonucleic acid (pre-mRNA) transcripts as they exit DNA polymerase II. As the pre-mRNA transcripts emerge from the polymerase complex, they are bound by hnRNP C only if its nucleotide (NT) chain is longer than a certain nucleotide length. If the chain is long enough and binding occurs, the nucleotide strand is exported and processed as mRNA, whereas if the length requirement is not met, the RNA sequence is directed along a pathway to become small nuclear RNA (snRNA). The functional form of hnRNP C is a four-copy tetramer, with four C proteins associating at their leucine zipper-like oligomerization domains (CLZ) to form the complex. The CLZ domains form a coiled-coil tetramer, a very stable structural motif in biological macromolecules. This work studies the thermodynamic properties of the CLZ domain of hnRNP C. Fluorescence resonance energy transfer (FRET) assays were conducted using fluorescently labeled CLZ peptides to determine the equilibrium dissociation constant (K_D) as well as the rate of tetramer formation (k_{on}). The experimental data was used in conjunction with computational simulations to gain a more coherent picture of CLZ interactions.

Acknowledgements

I would like to acknowledge the major efforts given by Dr. Kristin Whitson on this project. Without her mentorship, guidance, and inspiring passion for science this project would never have been finished. I want to acknowledge our collaborators at Vanderbilt University, Dr. Carlos Lopez and James Pino, for their work on the computational facet of this project. I also would like to acknowledge the efforts of Bomi Kim and Dr. Stephanie R. Whitson for help on the project as well. A special thanks also is needed for the Provost Student Research Awards for funding this project.

Contents

List of Figures

List of Tables

Abbreviations

bZLM	**b**asic region **Z**ipper-**l**ike **m**otif
CD	**C**ircular **D**ichroism
CLZ	hnRNP **C** **L**eucine **Z**ipper Oligomerization Domain
CTD	**C**-**T**erminal **D**omain
DNA	**D**eoxyribo**n**ucleic **A**cid
EMR	**E**lectro**m**agnetic **R**adiation
FRET	**F**luorescence **R**esonance **E**nergy **T**ransfer
F-CLZ	**F**luorescein labeled **CLZ**
hnRNP C	**h**eterogeneous **n**uclear **R**ibo**n**ucleo**p**rotein **C**
IR	**I**nfra**r**ed
kDa	**k**ilo**Da**ltons
m7G	**7**-**m**ethyl **G**uanosine
mRNA	**m**essenger **R**ibo**n**ucleic **Acid**
NMR	**N**uclear **M**agnetic **R**esonance Spectroscopy
NT	**N**ucleo**t**ide
pol II	**pol**ymerase **II**
RNA	**R**ibo**n**ucleic **Acid**
RRM	**R**NA **R**ecognition **M**otif
snRNA	**s**mall **n**uclear **R**ibo**n**ucleic **Acid**
T-CLZ	**T**etramethylrhodamine labeled **CLZ**
TMR	**T**etra**m**ethyl**r**hodamine
U-CLZ	**U**nlabeled **CLZ**

Physical Constants

Speed of Light	c	=	$2.997\ 924\ 58 \times 10^{8}\ \mathrm{ms}^{-1}$ (vacuum)
Gas Constant	R	=	$8.314\ 51\ JK^{-1}mol^{-1}$
		=	$1.987\ 20 \times 10^{-3}\ kcalK^{-1}mol^{-1}$
Avogadro's Number	N	=	$6.022\ 137 \times 10^{23}\ mol^{-1}$

Symbols

λ	wavelength	meters
n	refractive index	unitless
κ	dipole-dipole orientation factor	unitless

This thesis is dedicated to my UTC family, which includes faculty in both Physics and Chemistry, friends, and the UTC Cross Country and Track & Field team. You all have been instrumental in shaping me into the person I am today.

Chapter 1

Introduction and Background

1.1 hnRNP C

One of the greatest achievements of cell division is the high conservation rate of the genetic code. With the requirement for extreme accuracy accompanying the large number of steps in copying, transcribing, and translating the genetic code, there is much room for awe in the process's success. However, the success of this process is no accident. Synchronized efforts of multiple macromolecules guide the process, with each macromolecule working by itself or in collaboration with others to provide support, stability, and direction. The seamless activity often resembles a well-oiled machine, where the parts are the biomolecules associating and dissociating to achieve tasks. This thesis investigates the nature of a key aid in the process of cell division: heterogeneous nuclear ribonucleoprotein C (hnRNP C). Specifically, it aims to elucidate the thermodynamic stability of the CLZ oligomerization domain of hnRNP C proteins.

During transcription, RNA polymerases I, II, and III transcribe the genetic code from deoxyribonucleic acid (DNA) to synthesize precursors to various ribonucleic acid (RNA) molecules [1]. After these precursors are transcribed, several post-transcriptional processes occur to modify the transcripts into functional forms. Some of these processes include 5' capping, 3' cleavage and polyadenylation, pre-mRNA splicing, and export from the nucleus [1, 2]. Polymerase II (pol II) primarily synthesizes precursors to messenger RNAs (mRNAs) and some small nuclear RNAs (snRNAs). mRNAs carry information from the DNA code to the ribosome for which protein to synthesize, while snRNAs typically perform roles in RNA processing such as splicing [1]. The process from initial transcription by pol II to mature mRNA or snRNA involves several proteins and enzymes working in tandem.

hnRNP C belongs to a family of around 40 hnRNP molecules, labeled from A to U, that range in molecular mass from 32-120 kiloDaltons (kDa) [2, 3]. hnRNPs perform several nonredundant functions in the cell ranging from cytoplasmic mRNA translation and turnover to transcription and pre-mRNA processing [2, 4]. There are generally two classes of hnRNPs: those that leave the nucleus and those that are retained within the nucleus. Those that possess nucleocytoplasmic transport properties are hnRNP A1, K, and E [5]. Their ability to transport across the nuclear membrane is thought to originate from nuclear export signals that the hnRNP possesses [5]. hnRNPs C1, C2, and U are thought to override these export signals by a nuclear retention sequence of 78 amino acids not present in A1, K, and E [5].

Heterogeneous nuclear ribonucleoprotein C performs a critical role in processing

early strands and determining which strands become pre-mRNA or snRNA [6]. As pol II begins transcribing a new strand of RNA, a capping enzyme adds a 7-methyl guanosine (m7G) nucleotide (NT) to the 5' end of the growing strand through a 5'-5' linkage [1, 6]. Usual linkages are performed in a 5'-3' direction, and this unusual linkage inhibits degradation and thus aids in protecting the transcript as it grows. As the nascent transcript continues to grow, if it passes around 200 to 300 NTs in length it will associate with hnRNP C, which will begin the transcript on the pathway for export as mRNA. After the initial association of hnRNP C with the RNA, other hnRNP particles associate with the hnRNP-pre-mRNA complex to form a particle, which sends the RNA further down the processing pathway. If the NT length requirement is not met, however, hnRNP C doesn't associate with the transcript and the strand is marked for the snRNA export pathway [6].

The functional hnRNP C complex is composed of four associating C proteins. There are two types of C proteins found in the nucleus (C1 and C2) that are very similar and only differ by a 13 amino acid addition at Glycine 106 [2, 7–9]. In the nucleus, C1 and C2 form a 3:1 heterotetramer ($C1_3C2$) [2, 6, 10]. However, homotetramers of $C1_4$ and $C2_4$ have also been found to naturally occur as well [2, 11, 12]. Each type of C protein contains four functional domains that are highly conserved, the RNA recognition motif (RRM), a basic region zipper-like motif (bZLM), a leucine zipper-like oligomerization domain (CLZ), and an acidic C-terminal domain (CTD). The RRM (residues 8-87) and the bZLM (residues 140-179), are the key pre-mRNA binding regions of C proteins. Figure 1.1 presents a schematic of the hnRNP C sequence with the RRM, bZLM, CLZ and CTD regions labeled. CLZ (residues 180-207) is responsible for the organization of

C proteins into the functional hnRNP C tetramer, and the CTD (residues 208-290) functions in regulating some RNA-protein interactions [2]. Although CLZ oligomerization has been identified as limiting step in hnRNP C formation, little work has been done to elucidate the thermodynamic properties of the process and the tetramer itself.

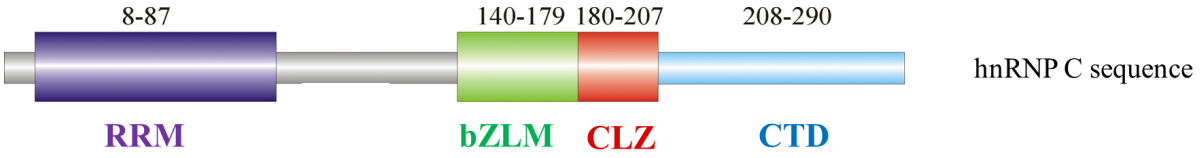

FIGURE 1.1: hnRNP C sequence.

With the first step in the pre-mRNA export pathway relying on functional hnRNP C, and hnRNP C functionality being derived from interactions of the CLZ region of C proteins, this thesis investigates the stability of the CLZ oligomerization domain of hnRNP C proteins.

1.2 hnRNP C Oligomerization Domain: CLZ

In the nucleus, the association of four C protein oligomerization domains (CLZ) is a key event in the formation of the larger hnRNP C complex. Each CLZ domain is a 28 residue region with the amino acid sequence LQAIKKELTQIKQKVDSLLENLEKIEKE [2]. The solution structure of an hnRNP tetramer is shown in Figure 1.2.

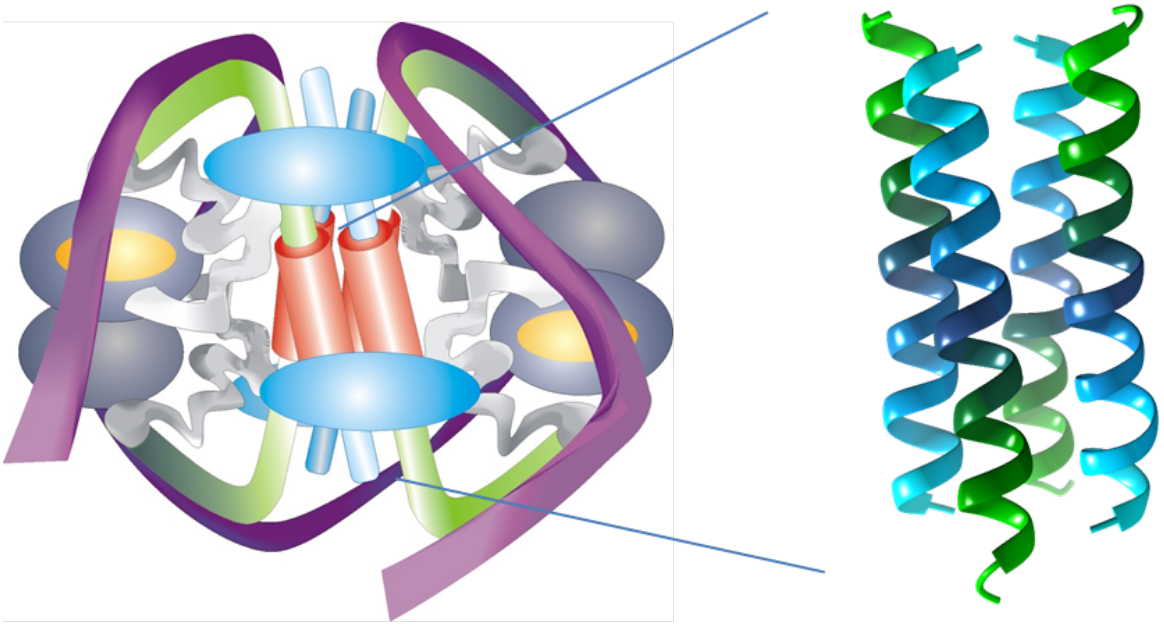

FIGURE 1.2: CLZ structure from protein databank ID 1TXP, and overall hnRNP C structure and organization as proposed by Whitson (2005)[2].

The amino acid sequence ultimately drives the types of secondary structure a protein forms, with each amino acid's polarity, charge, size, and hydrophilic/hydrophobic nature leading towards certain energy minimized secondary structures. Typically, proteins associate with themselves or other proteins to adopt structures that hide hydrophobic sidechains from aqueous solvent exterior (hydrophobic packing). This exposes polar and charged hydrophilic residues to the outside of the protein structure. Several common secondary structures that are frequently found in proteins effectively minimize the energies of the molten globular state of a protein. One of the most common structures is the alpha-helix [13, 14]. The alpha helix is characterized by phi and psi angle pairs of $-60°$ and $-50°$, respectively [2, 14–16]. These dihedral angles produce 3.6 residues per turn and rise per residue values of around 1.5 Angstroms. The alpha helix stabilizes the structure by allowing hydrogen bonds to occur between the carbonyl oxygen of residue

n and the amino group of residue $n + 3$ [14, 16]. While alpha-helices are very stable secondary structures by themselves, certain tertiary and quaternary structure formations can be formed that further minimize the energy of the molten globule state. Alpha-helices in particular can form coiled-coils with up to 4 other alpha-helices, producing supercoils. These interactions further bury hydrophobic residues through contacts with surrounding amphipathic alpha-helices with knobs-into-holes packing [2]. When the individual helices associate, they are either parallel or antiparallel. Parallel refers to their N-terminus and C-terminus ends being aligned on the same side of the coiled coil, whereas anti-parallel coiled coils have the N-terminus of one alpha-helix at the same end of the other helix's C-terminus.

Nuclear magnetic resonance (NMR) spectroscopy and circular dichroism (CD) spectroscopy experiments have shown CLZ forms alpha-helical secondary structures *in vitro* [17, 18], and further experiments have shown that these helices associate to form anti-parallel coiled-coil tetramers [17]. A representation of the CLZ anti-parallel coiled-coil tetramer is shown in Figure 1.3. Tetramerization into coiled coil helices is a stable and common structural motif. DNA-binding proteins such as the lactose operon (Lac) repressor [19] and tumor suppressor p53 [20] are two examples of evolutionarily conserved proteins that utilize four-helix assemblies for function. The antiparallel coiled coil formed by the CLZ helices has been found to be extremely stable, resisting dissociation at temperatures up to $89°C$ experimentally [18].

A consequence (or a driving force) behind the anti-parallel orientation of the CLZ helices is the juxtaposition of the other functional domains of C protein. The RRM

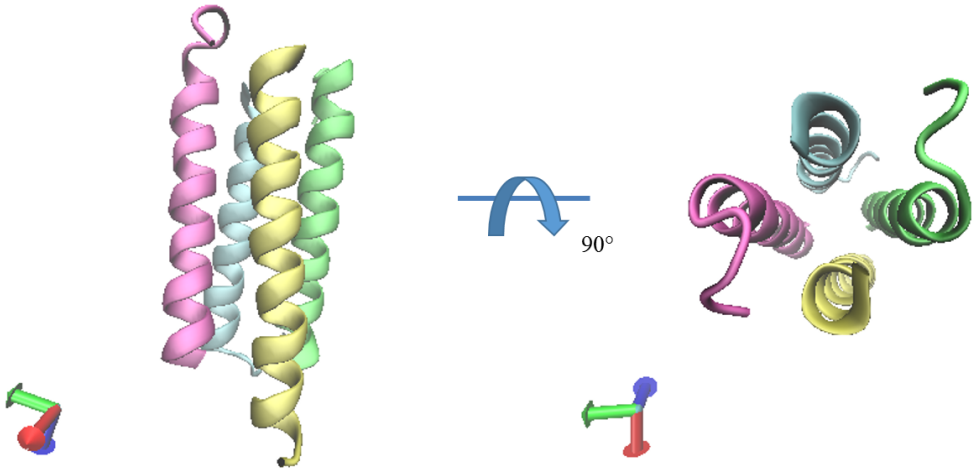

FIGURE 1.3: CLZ alpha helix structure from protein databank ID 1TXP.

and bZLM are both N-terminal, or upstream, in reference to CLZ, while the CTD is C-terminal, or downstream. With the anti-parallel orientation, the RNA-binding regions (RRM and bZLM) are more spatially separated, allowing for wrapping of longer RNA transcripts [2].

The stability of the CLZ coiled coil structural motif is not surprising, as a high degree of favorability would be expected for such an important process in cell division. In thermodynamic terms, this would mean an overall spontaneous process, and correspondingly, a very negative value for the standard Gibbs free energy, ΔG°.

1.3 Principles of Fluorescence Spectroscopy

1.3.1 Fluorescence Background

When electromagnetic radiation (EMR) is incident upon a molecule several types of interactions are possible. The interaction that occurs is dependent upon the energy content of the EMR and the energy levels of the molecules. High energy EMR (x-rays and γ-rays) can contain enough energy to scatter electrons [21, 22]. For some molecular species, EMR in the visible region matches the energy gap between molecular energy levels and an electron is promoted to an excited state [23]. Infrared (IR) spectroscopy studies the interaction of EMR with lower energy than the visible region (IR and radio) that induce vibrational or rotational molecular excitation. Fluorescence deals with EMR with energies that match the gap between molecular energy levels, typically in the visible wavelength region of the EMR spectrum [16].

Molecular species have certain degrees of freedom for electronic levels, vibrational levels, and rotational levels. Each electronic energy level has associated vibrational and rotational levels. Each combination of electronic, vibrational, and rotational energy levels represents a distinct molecular energy and are separated by discrete energy differences. While each type of level has discrete energy differences between adjacent levels, all levels are not created equal; the energy gap between electronic levels is significantly greater than that of vibrational levels, and vibrational energy differences are greater than rotational. The lowest energy electronic state is defined as the ground level. At physiologically

relevant temperatures ($37°C$), molecular energy fluctuations are limited to the rotational and vibrational energy levels of the ground electronic energy level.

An electron can be promoted to an excited state if the incident photon energy matches the gap between the excited state and the initial ground state. After excitation, there are several deactivation pathways for returning to the ground state. Of these deactivation pathways, two emit photons: fluorescence and phosphorescence. Fluorescence is the emission of photons from the return of an electron to the ground electronic state [16]. For fluorescence, an electron's lifetime in the excited state is around 10^{-9} seconds [23]. Phosphorescence involves an excited electron switching spin states from a singlet state to a triplet state before returning to the ground state. The process of switching from a singlet state to a triplet state is not very probable, and can take from 10^{-4} to 10 seconds to occur [23].

Fluorescence and phosphorescence are radiative deactivation pathways. However, there are many deactivation pathways that do not emit photons. Some of the most pertinent mechanisms include external conversion, internal conversion, intersystem crossing, predissociation, and dissociation [23]. External conversion refers to the loss of energy through collisions with solvent molecules or other molecules in solution. Internal conversions are intermolecular interactions that require energy, such as switching two states of the same multiplicity. Intersystem crossing involves changing multiplicity, for example the conversion from singlet to a triplet state. Dissociation is the deactivation pathway where a molecular bond is broken. The preferred deactivation pathway for an excited molecule is the one in which the excited lifetime is minimized. Said another way, the

process with the fastest rate will be the most favorable.

The rate of fluorescence emission can be expressed as the quantum yield (Q), or probability flux. The quantum yield for emission can be represented by the Einstein relation in Equation 1.1 [16]. In Equation 1.1, T is the probability that the excited molecule loses a quantum of energy by any process and A is the Einstein coefficient that describes the probability of spontaneous emission.

$$Q = \frac{A}{T} \tag{1.1}$$

Another way to write the quantum yield is illustrated in Equation 1.2. In Equation 1.2, k_f represents the rate of fluorescence, k_{ec} represents the rate of external conversion, k_i represents internal conversions, k_{ic} represents intersystem crossing, k_d represents the rate of dissociation, and k_p represents the rate of phosphorescence. The rates k_d, k_f, and k_p are structure dependent, whereas the other rates are influenced by environmental factors [23, 24].

$$Q = \frac{k_f}{k_f + k_{ec} + k_i + k_{ic} + k_d + k_p} \tag{1.2}$$

A simpler representation is given by Equations 1.3 and 1.4, where the sum of all the non-fluorescent rates are represented by tau (τ).

$$\tau = k_{ec} + k_i + k_{ic} + k_d + k_p \tag{1.3}$$

$$Q = \frac{k_f}{k_f + \tau} \tag{1.4}$$

If $k_f >> \tau$, fluorescence will be a dominant deactivation pathway and the probability of fluorescence occurring is high. If $k_f << \tau$, the probability flux of fluorescence will be very small. Possibly the biggest determinant for the rate of fluorescence is molecular structure. Certain molecules with resonance structures have lower molecular energy gaps that are favorable for fluorescent transitions. With that in mind, only certain molecular structural groups, called fluorophores, can be used for fluorescent experiments. For many biomolecules, fluorophores are covalently linked to non-fluorescent molecules so they can be analyzed [25].

Despite requirements of discrete energy amounts for excitation and emission transitions, excitation and emission are not found at only single-wavelength values. The broad bands for both excitation and emission are results of the vibrational and rotational energy levels associated with each electronic level that are available for transitions [23, 24].

Typical fluorescence instrumentation uses visible light to excite a sample and the fluorescence is observed at right angles. There is a monochromator before the sample to select the wavelength of exciting light, and a monochromator after the sample selects the wavelength of emitted light to observe [16, 23]. Two types of scans are conducted for fluorescence experiments: excitation and emission scans. Excitation scans set a fluorimeter to record emission intensity at a single wavelength, while the incident light wavelength is changed by turning a diffraction grating (the excitation monochromator) in the instrument. Excitation spectra show the incident wavelength (energy) that produces the

greatest emission of photons. Emission scans use a single incident wavelength and the detector scans a range of wavelengths. Fluorescence emission spectra show the range of wavelengths that are emitted when a molecular species is excited by a certain wavelength of light.

1.3.2 Fluorescence Resonance Energy Transfer

Fluorescence resonance energy transfer (FRET) is an alternative deactivation pathway to fluorescence. The process involves two fluorophores, a donor and acceptor [24]. When the donor fluorophore is excited, the energy can nonradiatively excite the acceptor molecule, which can then emit its distinct spectrum. There are several key factors that influence the rate of fluorescence, including: spectral overlap between donor emission and acceptor absorption, distance between the fluorophores, and the alignment of the fluorophore transition dipole moments.

The spectral overlap is specific for each donor-acceptor pair and can be represented by the Forster distance (R_0). The Forster distance is defined as the distance apart fluorophores are such that FRET occurs with 50% efficiency. An expression for the Forster distance is given in Equation 1.5.

$$R_0^6 = (\frac{9000(ln10)\kappa^2 Q_D}{128\pi^5 N n^4})J \tag{1.5}$$

In Equation 1.5, κ^2 is the dipole-dipole orientation factor, Q_D is the quantum yield of the donor fluorophore, N is Avogadro's number, n is the refractive index of the medium, and

J is the overlap integral between the donor emission and acceptor excitation. Equation 1.6 presents the full expression for J.

$$J = \int_0^\infty F_D(\lambda)\epsilon_A(\lambda)\lambda^4 d\lambda \tag{1.6}$$

In Equation 1.6, $F_D(\lambda)$ represents the fluorescence intensity of the donor fluorophore and $\epsilon_A(\lambda)$ represents the extinction coefficient of the acceptor as a function of wavelength (λ). Figure 1.4 presents the overlap integral for F-CLZ and T-CLZ.

The relationship between FRET efficiency (E), the Forster distance (R_0), and the distance between the donor and acceptor fluorophores (r) is shown in Equation 1.7 [16, 24, 27]. Note that FRET efficiency decreases as an r to the sixth power, so extremely small distance deviations can be detected. In fact, FRET is commonly called the spectroscopic ruler [28, 29].

$$E = \frac{R_0^6}{R_0^6 + r^6} = \frac{1}{1 + (\frac{r}{R_0})^6} \tag{1.7}$$

FRET also has a requirement for the transition dipole moments of the two fluorophores to be aligned. The dipole orientation factor, κ, is represented in Equation 1.5. The degree of dipole orientation ranges from 0 (not aligned) to 1 (perfectly aligned). κ^2 is assumed to be $\frac{2}{3}$ for calculations of Forster distances [24], which is appropriate for dynamic random averaging of the orientations of the molecules in solution.

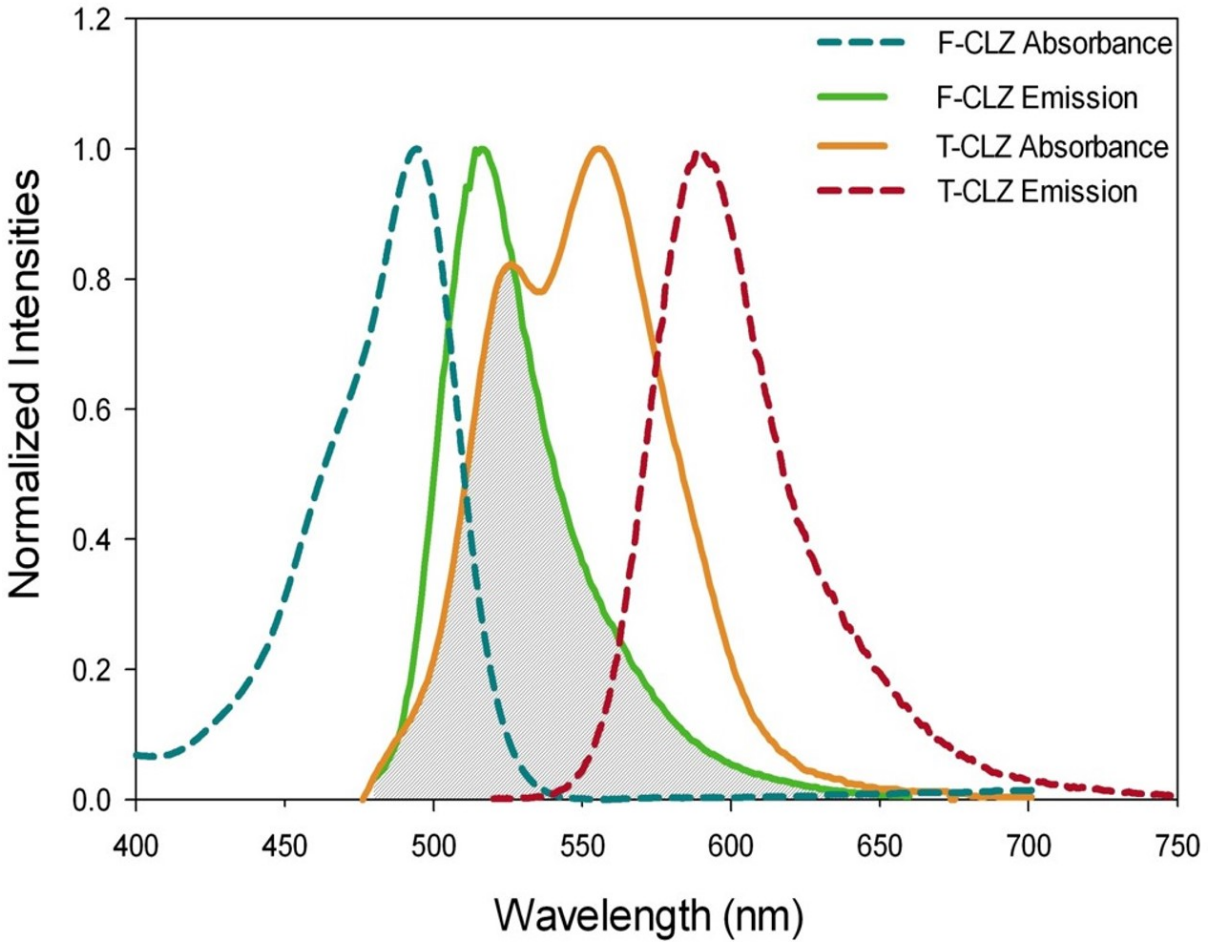

FIGURE 1.4: Overlap integral $J(\lambda)$ for F-CLZ and T-CLZ (Gray region). Adapted from Pino *et al*[26].

There are many uses for FRET in biophysical research. Two of the most common uses are for investigating distance relationships and associations of biological macromolecules. Lakowicz *et al.* (1990) have used FRET to measure distances between amino acids in alpha-helical melittin [30]. Protein folding studies on serine hydroxymethyltransferase have utilized FRET as well. Three tryptophan residues were labeled for FRET analysis and were used to determine which regions of the polypeptide come together first [31]. Ghosh *et al.* (1994) used FRET to study the binding of PaeR7 endocuclease to

DNA [32]. With both the DNA strand and PaeR7 labeled with fluorophores, upon binding FRET was observed. When the molecules dissociated, the emission spectra of the donor returned to its normal shape.

1.4 Project Goals

This project seeks to investigate the CLZ oligomerization process both experimentally and computationally. There are several models for which the oligomerization process can occur, and with both experimental and computational methods having different areas of applicability, we aim to provide a coherent picture of the process by combining the approaches.

All that is currently known about CLZ oligomerization is the overall net mechanism presented below, when the reaction is at equilibrium. Equation 1.8 presents the relationship between the rate of formation of the tetramer complex (k_{on}) and the rate of dissociation of the tetramer into monomers (k_{off}); the ratio of the two processes is defined as the equilibrium dissociation constant, K_D.

$$4 \text{ Monomers} \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} \text{Tetramer}$$

$$K_D = \frac{k_{off}}{k_{on}} = \frac{1}{K_A} \tag{1.8}$$

The goal of this thesis is to provide further insight into the tetramerization mechanism. We propose that the mechanism of tetramer formation can be illustrated through

a series of elementary reactions. Elementary reactions are characterized by the number of reactants, and are typically defined as unimolecular for single molecule reactions, bimolecular for two molecule reactions, and termolecular for three molecule reactions [27]. We propose a mechanism involving two reversible bimolecular events: the association of two monomers to form a dimer, and the subsequent association of two dimers to form the tetramer:

$$\text{Monomer} + \text{Monomer} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \text{Dimer}$$

$$\text{Dimer} + \text{Dimer} \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} \text{Tetramer}$$

Equilibrium and kinetic experiments are both used to elucidate the stability of the complex and its assembly rate. Quantitatively determining the equilibrium dissociation constant (K_D) was the primary goal for equilibrium experiments. The K_D can be related to stability of tetramer formation through Equation 1.9. In Equation 1.9, ΔG° is the standard-state free energy change, R is the Gas constant, T is the temperature for the system, and K_D is the dissociation constant.

$$\Delta G^\circ = -RTln(\frac{1}{K_D}) \tag{1.9}$$

The dissociation constant provides no insight into the rate of tetramer assembly. Kinetic experiments were done to model the rate of tetramer formation and find the rate

of formation k_{on}. The equilibrium and kinetic experimental data will be used to train computational simulations of the process, which will then allow for extrapolation of k_{on}, k_{off}, and K_D at different temperatures using advanced simulation techniques.

Chapter 2

Materials and Methods

2.1 Materials

The 25-mer CLZ peptides were obtained from the Genscript peptide synthesis service (Piscataway, NJ). The amino acid sequence was was IQAIKKELTQIKQKVDSLLENLEKI. Fluorophores were covalently attached to the N-terminus of the peptide. The F-CLZ peptide is modified by attachment of fluorescein isothiocyanate with an amino-hexanoic acid spacer, which is a 6-chain carbon linker, to the N-terminus of the protein. The T-CLZ peptide has the tetramethylrhodamine fluorescent moiety directly attached at the N-terminus via a covalent bond. The U-CLZ peptide is unlabeled by a fluorophore.

Standard solutions of T-CLZ, F-CLZ, and U-CLZ were made from the $1mg$ peptide stocks from Genscript. The three standards were all prepared in the same manner. The proteins were reconstituted in 1,500 μL of 50 mM Tris buffer or 20 mM HEPES buffer

that was prepared using MilliQ purified water (Millipore, Billerica, MA) and pH adjusted to 7.4 with NaOH. The solution was then vortexed before being placed in a water bath at 37 $°C$ for ten minutes to solubilize the protein. The protein solutions were then stored at -20$°C$ or -80$°C$ to maintain stability.

Prior to experiments, the protein solutions were placed in a 37 $°C$ water bath for between 1 and 48 hours to equilibrate the solutions to the experimental temperature.

2.2 Experimental Parameters

A Qantamaster 40 fluorimeter (Photon Technology, International, now Horiba, New Jersey, USA) and a 3x3x40 mm fused silica micro-cuvette were used for all fluorescence experiments. The equipment is temperature controlled by means of an external circulating water bath that maintains the temperature of the sample chamber. Several instrumental parameters will be explained briefly here with their impact on experimental data as they are discussed further below in the experimental methods sections. The excitation wavelength refers to the wavelength that the diffraction grating in the monochromator prior to the sample selects. For excitation scans, the diffraction grating rotates to focus a different wavelength of light on the sample. In emission scans, the excitation wavelength is fixed, while the diffraction grating in the emission monochromator before the photomultiplier tube detector rotates to scan through a range. Between the excitation monochromator and the sample chamber is an excitation slit that effectively decreases the intensity of light that is incident upon the sample. A smaller slit width reduces the incident intensity

upon the sample, but there is another effect from reducing the slit width besides intensity change. The slit size also controls what frequencies (or wavelengths) are incident on the sample. Thus, a 1 mm slit width essentially serves as a 4 nm bandpass filter, while a 2 mm slit width is excited by light in a 8 nm bandpass around the centrally designated excitation peak wavelength. At a 90°angle to the incident light is an emission slit that determines the intensity of emitted light that reaches the detector as well as the range of wavelengths detected.

Apart from the optical aspects of the fluorimeter, the experimenter also has control over the time for the excitation and emission scans. The integration step size determines the wavelength increments that the fluorimeter scans through, and the integration time is the length of time the photomultiplier tubes collect photons at a respective wavelength. A longer integration time allows for a higher number of emitted photons to be detected for each wavelength, which reduces random error and thermal noise fluctuations and allows for a more accurate value of the intensity to be determined at each wavelength. The cost of the increased information about the spectra is the time required for one complete scan. In time-sensitive experiments such as our kinetic rate experiments discussed below, scans are needed rapidly to produce an accurate profile of the rate mechanism. The exact specifications needed for our experiments could only be found through specific tests to see which produced the spectra better-suited for our goals.

For example, we first ran fluorescence emission experiments to try to observe FRET through traditional fluorescence emission scans. In these experiments, the solutions were excited using 468 nm light, and the detector scanned emission intensity from 478-725

nm. The integration time for each wavelength was set to 0.5 seconds, and the slit widths leading to the sample cuvette and to the detector were set to 0.5 mm.

Another test was run to observe the optimal load volume for measurements. Previously the group had used 200 μL solutions. The load volume was tested for 200 μL and 250 μL and the resulting curves were found to be the same, inferring 200 μL provided enough height in the cuvette for all of the sample to be within the beam path.

2.3 Kinetic Experiment Method

Solutions of equal concentrations of F-CLZ and T-CLZ were mixed together and excitation spectra were taken using the Quantamaster 40 fluorimeter at specific time points to measure the appearance of FRET. When the two solutions are mixed, they initially exist as homotetramers of solely F-CLZ and T-CLZ, assuming the concentration is greater than the equilibrium dissociation constant, K_D. If the concentration is below the K_D value, the solutions consist primarily as the subunits of the tetramer. At equilibrium, however, the homogeneous mixture would be expected to contain tetramers with an even distribution of both T-CLZ and F-CLZ. This assumes that the fluorescent labels themselves do not affect CLZ tetramer binding affinity. Said alternatively, upon reaching equilibrium the tetramers exist in 2:2 ratio of F-CLZ:T-CLZ unless there are different affinities between T-CLZ:T-CLZ helices or F-CLZ:F-CLZ helices. As the F-CLZ and T-CLZ monomers associate to form the coiled coil tetramers, FRET is observed between the donor fluorescein tag and the acceptor tetramethylrhodamine tag.

The fluorescence spectra were produced by monitoring the emission at 650 nm, a wavelength where it is expected that tetramethylrhodamine molecules are primary contributors to the emission. The incident light was scanned to excite both fluoroscein and tetramethylrhodamine in the wavelength range of 400-640 nm. In these experiments, we expect that the efficiency and intensity of FRET increases until equilibrium is reached, at which point the FRET ratio will remain constant with regards to time. Another term used to describe FRET is quenching, where it is inferred that the acceptor molecule "quenches" the fluorescence of the donor. Thus, FRET can be monitored by the appearance of the acceptor's fluorescence or by the corresponding quench of the donor's fluorescence intensity. Therefore, fitting the FRET ratio or the integrated intensity of either the donor or acceptor as a function of time allows for the determination of k_{on}. The model of tetramer formation was evaluated experimentally by monitoring the appearance of FRET as a function of time.

Kinetic experiments were conducted at a temperature of $37°C$ for several concentrations of CLZ to determine the consequences of concentration on the rate of tetramer formation. Concentrations of 500 nM, 1 μM, and 30 μM were investigated. Solutions of each concentration with either T-CLZ or F-CLZ were mixed at time zero and vortexed for 5 seconds to ensure the solutions themselves were mixed together before 200 μL was pipetted into the fluorimeter cuvette. After the cuvette was placed inside the fluorimeter the first scan was acquired. While the time between mixing and the first measurement varied slightly with each concentration, it typically was around 45 seconds. Ideally, scans would be acquired quickly and in a back-to-back fashion so as to attain the clearest picture of the kinetic formation of the tetramer. However, as the scan rate increases, peak

resolution decreases, and continuous scanning back-to-back can cause photobleaching of the fluorophores in the sample. Photobleaching is the phenomena where a fluorophore is oversaturated with incident excitations so the molecules become stuck in their excited state and emission intensity decreases as a result [30]. For these reasons, scans were acquired every 2.5 minutes for the first 45 minutes, and every 10 minutes after that until changes were no longer observed.

2.4 Equilibrium Experiment Method

Table 2.1 presents the concentrations used for the equilibrium experiments, as well as the emission slit width settings for each fluorimeter scan. The solutions were made by serial dilution, so as to minimize random error and conserve material. The solutions were placed in a water bath at $37°C$ for one to two days prior to measurements to ensure the solutions were at equilibrium. Scans were collected several days later to ensure that the low concentration samples had reached equilibrium. The equilibrium experiment was conducted at $37°C$ four times to provide confidence in the analysis.

The wide range of concentrations presented an interesting situation for instrumental settings. Ideally, the same instrument parameters would be used for all measurements. For this experiment, however, when the wide slit widths and long integration times required for capturing the spectra of the low concentration samples were applied to the high concentration samples the detector was oversaturated with signal and spectral resolution was lost. To introduce the smallest source of error into the experiment, only the

TABLE 2.1: CLZ concentrations used in equilibrium FRET experiments.

Number	Concentration (μM)	Emission Slit Width (mm)
1	0.003	2.5
2	0.01	2.5
3	0.03	2.5
4	0.10	2.5
5	0.05	2.5
6	0.07	2.5
7	0.10	2.5
8	0.15	2.5
9	0.20	2.5
10	0.30	2.5
11	0.50	2.5
12	0.70	2.5
13	1.00	2.5
14	2.00	2.5
15	3.00	2.5
16	5.00	2.5
17	7.00	2.5
18	10.0	1
19	15.0	1
20	20.0	1
21	30.0	1
22	40.0	1

emission slit widths were changed for different concentrations, while the excitation slit widths were not varied. Samples with 10 μM CLZ and higher utilized 1 mm emission slit widths, while 2.5 mm emission slit widths were used for all samples containing 7 μM or less CLZ. The excitation wavelength was scanned through 400-640 nm, while the emission wavelength of the detector was set at 650 nm.

Control samples were also made using unlabeled-CLZ (U-CLZ). Two sets of controls for all the concentrations presented in Table 2.1 were made with either equal parts U-CLZ and F-CLZ or equal parts U-CLZ and T-CLZ.

Chapter 3

Results and Discussion

3.1 Equilibrium Experimental Results

The raw excitation scans acquired from the equilibrium experiments are shown in Figure 3.1. In these experiments, the fluorescence intensity of the mixed F-CLZ/T-CLZ sample at 650 nm was observed as a function of the excitation wavelength. As the total concentration of the fluorophores in solution was increased, the fluorescence emission intensity increased, in accord with what is expected from the Beer-Lambert Law $A = \epsilon lc$, where A is the absorbance, ϵ is the molar absorptivity, l is the pathlength, and c is the concentration of the molecular species. Since a higher absorbance of light by the sample occurs, a higher fluorescence emission is necessarily expected. The broadband peaks seen in the spectra correspond to quantized energy transitions to varying electronic, vibrational, and rotational excited states from the ground state that give rise to a 650 nm photon. By comparison with the excitation spectrum of a T-cLZ standard, the peak

centered at 555 nm corresponds to the direct excitation of tetramethylrhodamine that results in T-CLZ emission. The peak centered at 496 nm corresponds to the emission of a 650 nm photon from T-CLZ that was put in the excited state by a FRET interaction with the donor fluorescein. In order to better visualize the peak wavelengths and their relative intensities within each sample, this data was normalized to the same maximum intensity (presented in Figure 3.2). The low concentration samples have dominant excitation peaks at 555 nm, the excitation peak for tetramethylrhodamine. As the total concentration of CLZ in the sample increases, the intensity from excitation at 496 nm increases compared to 555 nm. The appearance of the 496 nm excitation peak, and corresponding decrease in the 555 nm excitation peak is indicative of the appearance of FRET. The inflection point from when the ratio of emission intensity from 496 nm excitation to emission intensity from 555 nm excitation determines the K_D.

To quantify the extent of tetrameric complex versus individual subunits in the sample, the proportion of FRET (due to tetramer formation) was determined. Appendix A presents further explanation for the favorability for subunits at low concentrations and the tetramer at high concentrations. The amount of FRET is calculated by taking the ratio of the fluorescence from T-CLZ due to its excitation to the fluorescence from T-CLZ due to excitation of F-CLZ as shown in Equation 3.1.

$$FRETRatio = \frac{I_{555nm}}{I_{496nm}} \tag{3.1}$$

For low concentration samples, where the spectra revealed the most intense fluorescence in the yellow portion of the visible light spectrum, the ratio is a relatively large fraction. In high concentration samples where FRET occurs due to formation of tetramers, the spectra shows the most intense fluorescence in the green; thus the ratio is a relatively low fraction.

This data was plotted as a function of the concentration of CLZ in the sample (Figure 3.3, inset). Alternatively, the ratio of intensities was plotted against the log of concentration, which better reveals the inflection points in the curve. This is presented in Figure 3.3.

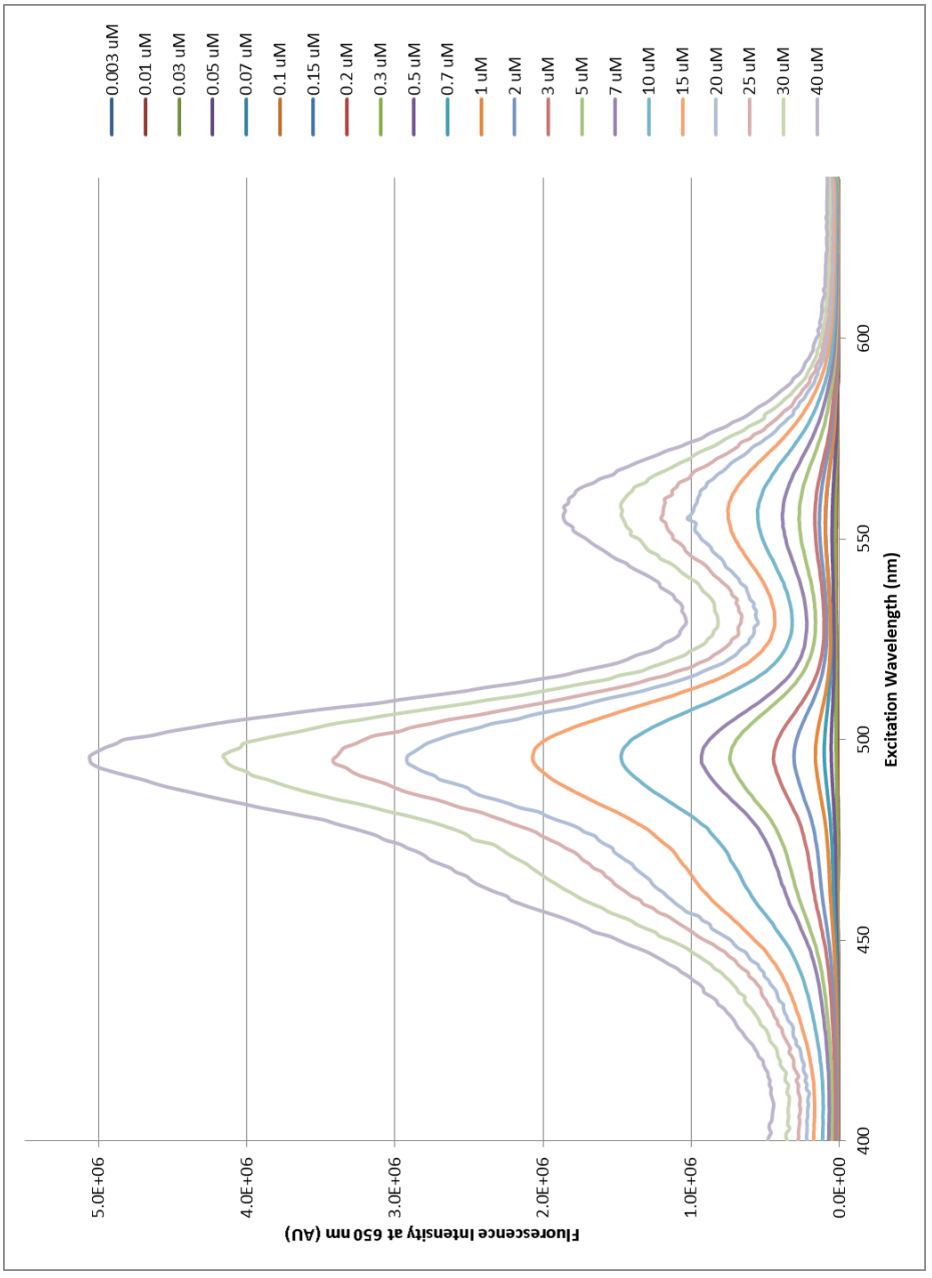

FIGURE 3.1: Data from a representative equilibrium binding experiment. The concentrations represent the total concentration of CLZ in the sample.

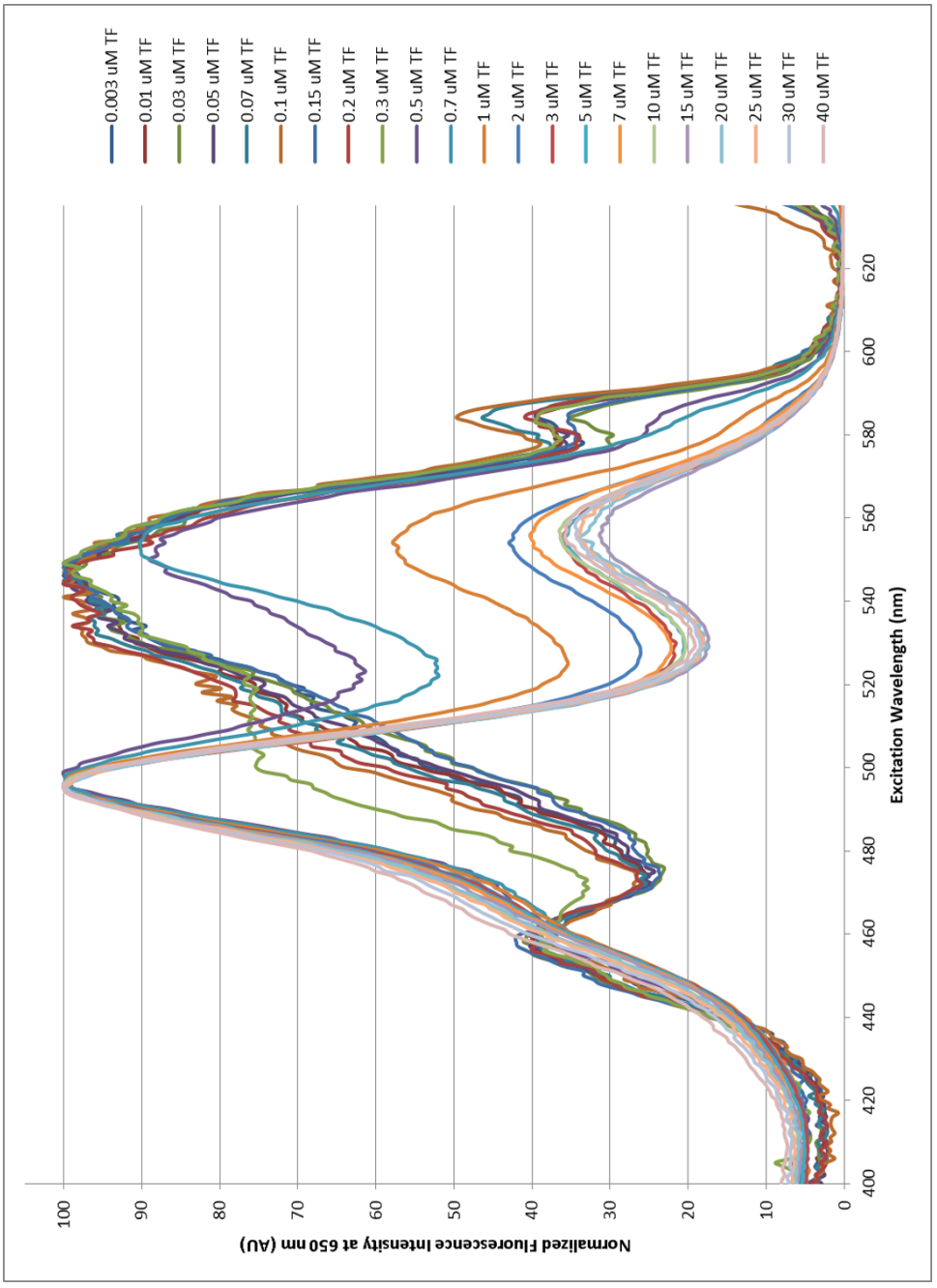

FIGURE 3.2: Normalized excitation spectra from the equilibrium experiment.

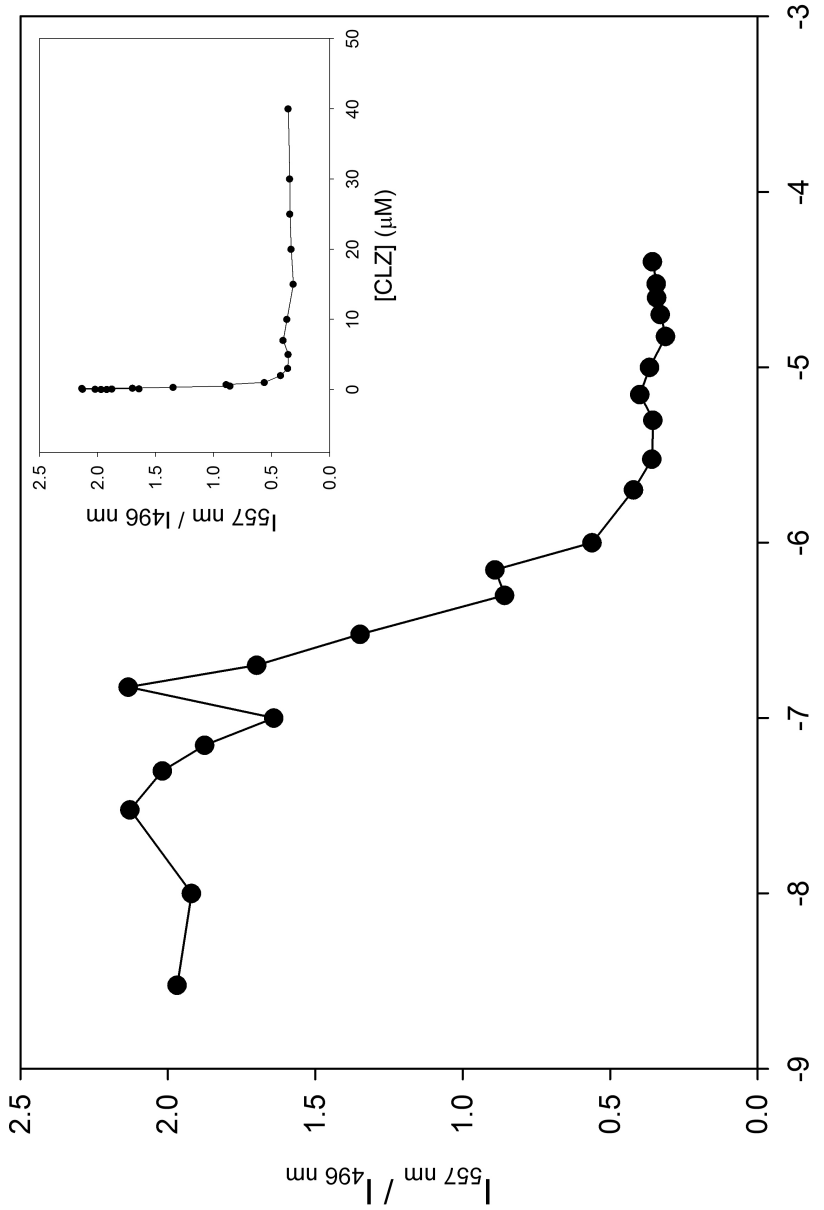

FIGURE 3.3: FRET ratio as a function of logarithmic CLZ concentration. (inset) FRET ratio as a function of CLZ concentration

The data follows the trend that we expect, except for the sharp decrease and subsequent increase in the ratio of intensities for concentrations around $100 - 250\ nM$. This anomaly is reproducible across multiple experiments and sample preparations. It is possible that it reflects a structural transition or a distinct step in the mechanism of tetramer formation.

The entire data set was fit to a four parameter logistic equation of the form presented in Equation 3.2.

$$y = D + \frac{A - D}{1 + 10^{B(x - log(C))}} \tag{3.2}$$

In Equation 3.2, A represents the maximum value of the ratio, D represents the minimum value of the ratio, C represents the inflection point at the half-maximum of the curve, and B represents the slope of the curve. The fit parameters and their values are given in Table 3.1, and the fit of the equation to the equilibrium data is shown in Figure 3.4.

TABLE 3.1: Fit parameters and values for four-parameter logistic fit.

$R^2 = 0.9773$			
Parameter	Coefficient	Standard Error	P
A	1.9898	0.0563	<0.0001
B	2.1235	0.3714	<0.0001
$log(C)$	-6.3949	0.0427	<0.0001
D	0.3518	0.0400	<0.0001

Alternatively, the data set was fit to the three-parameter logistic equation:

$$y = D + \frac{A - D}{1 + 10^{(x - log(C))}} \tag{3.3}$$

where A, C, and D represent the same variables as in the four-parameter fit (Equation 3.2), but the value for the slope, previously denoted as B, is forced to be equal to 1 in this case. This fit is shown in Figure 3.4 and the fit values are given in Table 3.2.

TABLE 3.2: Fit parameters and values for three-parameter logistic fit.

$R^2 = 0.9520$			
Parameter	Coefficient	Standard Error	P
A	2.1573	0.0942	<0.0001
$log(C)$	-6.4080	0.1016	<0.0001
D	0.02482	0.0644	<0.0001

The equilibrium dissociation constant, K_D, corresponds to the concentration at which half of the proteins are in the tetrameric complex that produces a FRET signal. Thus, the K_D is the parameter C in the logistic functions. The recovered value is $400 \pm 40\ nM$, with both fits yielding values that are within the error range of each other. The uncertainty in the K_D value was calculated using the standard error in $log(C)$ and propagation of error techniques.

While both fits give the same value for the K_D, the fit with the four-parameter logistic function is statistically better. Examination of the fits in Figure 3.4 reveals that the inflection points where the sigmoidal shape reaches the minimum and maximum values are missed by the three parameter function, even though the inflection point at the half-maximum of the curve (parameter C) is present.

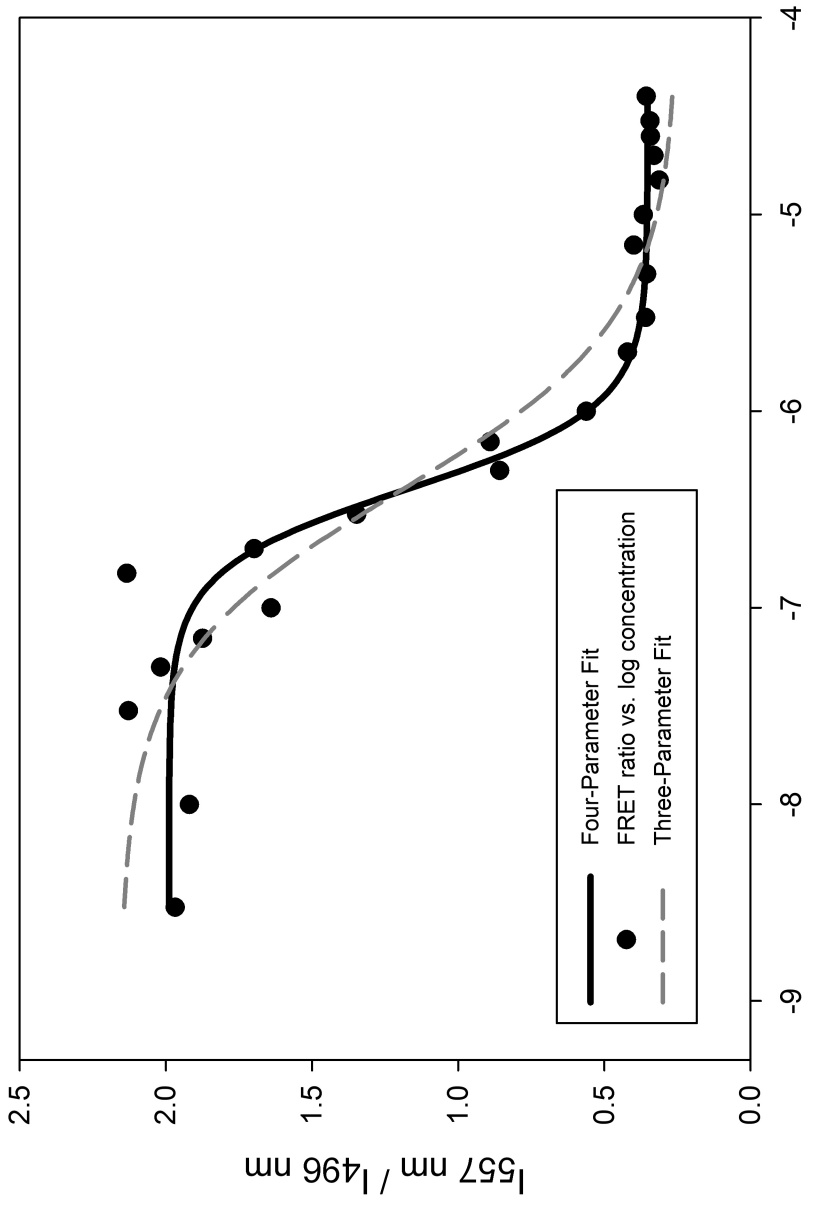

FIGURE 3.4: Equilibrium FRET ratio data fit with 3-parameter and 4-parameter logistic equations.

In ligand-binding models and theory, variable slope represented by B in the four-parameter equation has the physical significance of cooperativity in the system [33]. Positive cooperativity is indicated in the system if this slope is greater than 1, and would be where assembly of one FRET-binding complex aided the formation of another. Alternatively, if a multi-step process, such as formation of an intermediate dimer or the unfolding of a monomer into a proper orientation prior to assembly into a tetramer were required, this could be revealed in the data in the same way that positive cooperativity is shown.

From the equilibrium dissociation constant, we calculate the standard Gibb's free energy, ΔG°, for the formation of the tetramer by Equation 1.9. Using the temperature as $310K$, this yields a value of $-37.9712 \pm 0.0008\ kJ\ mol^{-1}$.

3.2 Kinetic Experimental Results

The kinetic experiment aimed to analyze the rate of FRET occurrence as a function of time after the mixing of two equal concentration portions of T-CLZ and F-CLZ. The kinetic data was fit to an exponential function of the form presented in Equation 3.4, and to the form presented in Equation 3.5.

$$y = y_0 + a(1 - e^{-bx}) \tag{3.4}$$

$$y = y_0 + a(1 - e^{-bx}) + c(1 - e^{-dx}) \tag{3.5}$$

Both fits were evaluated to investigate whether at this specific concentration (1 μM total CLZ), the association of the helices that result in FRET looked to be a one-step or two-step process. The fit to this data is shown in Figure 3.5, and the recovered parameters are shown in Table 3.3.

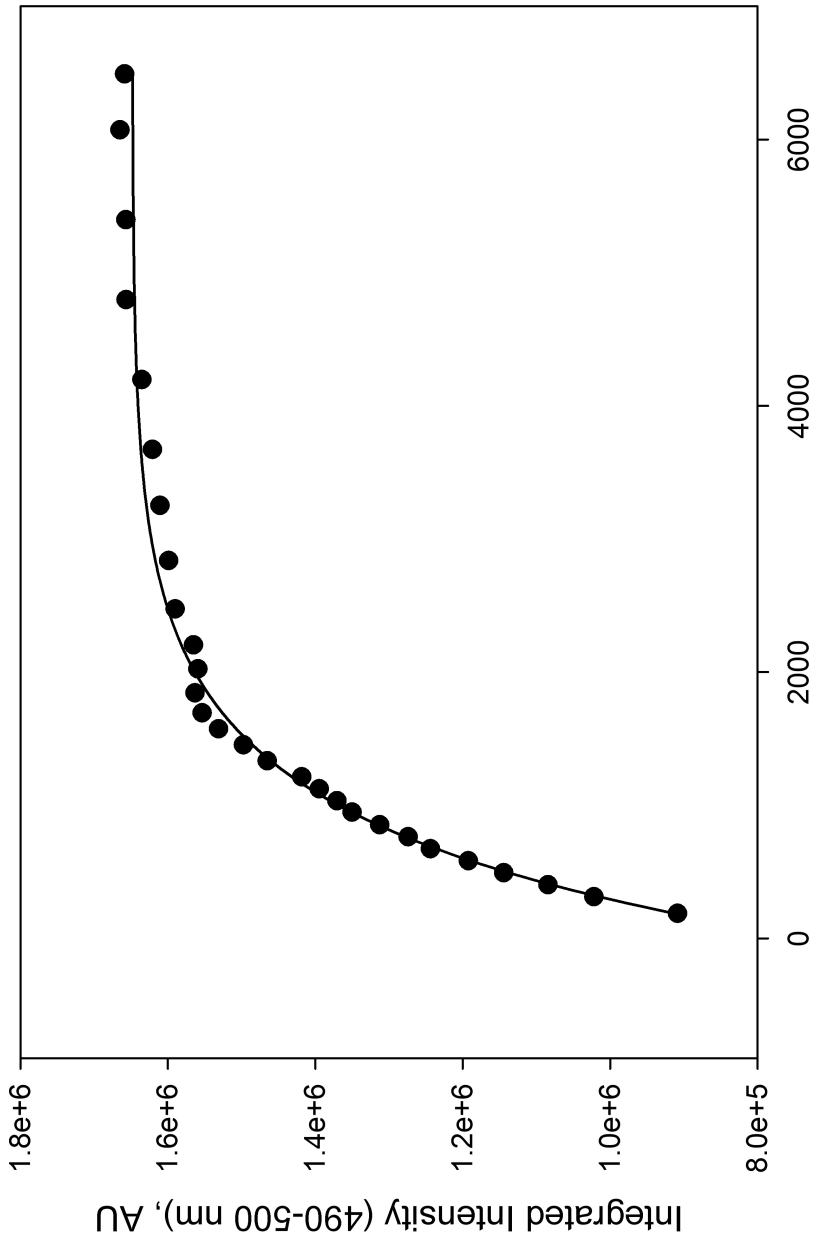

FIGURE 3.5: Exponential fit of kinetic FRET data.

The double-exponential function did not provide a better fit to the data and produced statistically irrelevant results for the second set of fit parameters. Therefore, the FRET signal that changed over this time-scale (roughly 45 minutes) for a 1 μM total CLZ concentration sample was due to only one process in the proposed mechanistic formation of the tetrameric species.

The value for the association rate for 1 μM CLZ is therefore $k_{on} = 0.00120 \pm 0.00003\ M^{-1}\ s^{-1}$. Using our determined value for the K_D we would predict a kinetic dissociation rate of $k_{off} = 4.8 \times 10^{-10} s^{-1}$. This is what we would expect for a highly stable complex; the subunits assemble rapidly, and would very infrequently dissociate from each other once the complex is made.

This experiment was also attempted using 30 μM total CLZ, but the change in FRET signal was too fast to observe well, and at 500 nM, the change was too slow, possibly because 500 nM is nearer to the K_D, so not enough tetramers were formed. The association rate, k_{on}, would be expected to vary for different concentrations, in the same manner observed. The dependence of this rate on the concentration would reveal whether this is a first, second, or fourth-order rate law for the kinetic assembly of the complex.

TABLE 3.3: Fit parameters and values for single and double exponential equations for rate of formation

Parameter	Single Exponential Fit	Double Exponential Fit
R^2	0.9964	0.9970 (p <0.0001)
y_0	$7.3 \times 10^5 \pm 0.1 \times 10^5$ (p <0.0001)	$7.1 \times 10^5 \pm 0.2 \times 10^5$ (p <0.0001)
a	$9.2 \times 10^5 \pm 0.1 \times 10^5$ (p <0.0001)	$9.0 \times 10^5 \pm 0.7 \times 10^5$ (p <0.0001)
$b(M^{-1}sec^{-1})$	0.00120 ± 0.00003 (p <0.0001)	0.0013 ± 0.0001 (p <0.0001)
c	-	$0.006 \times 10^9 \pm 4.244 \times 10^9$ (p = 0.9988)
$d(M^{-1}sec^{-1})$	-	0.000001 ± 0.000800 (p = 0.9988)

Chapter 4

Conclusion and Future Direction

4.1 Implications of Computational Fitting

Computational work on CLZ has been carried out in conjunction with collaborators at Vanderbilt University. All work was done using the same peptide sequence as our experimental work (protein data bank ID: 1TXP). Molecular dynamics simulations were conducted to calculate the standard Gibbs free energy of dissociation of both the dimer $\longrightarrow$ monomer and tetramer $\longrightarrow$ dimer transition. These simulations were run using the Gromacs molecular dynamics suite [34]. The simulations began by forming starting conformations of either a CLZ dimer or CLZ tetramer. After the energy of the complex was minimized, the center of mass between each subunit was defined. Each subunit was then simulated for 5 ns at 50 distances along the reaction coordinate extending from the center of mass. The force required to keep the center of masses apart at each distance along the center of mass coordinate was then calculated using the weighted

histogram analysis method [35]. This method recovers the standard Gibbs free energy of the complex.

The standard Gibbs free energy was used to calculate a K_D, and was also used to predict the model of tetramer formation. Our experimental equilibrium and kinetic data were used to parametrize the computational models of the tetramerization process. A set of nonlinear ordinary differential equations were created using the BioNetGen[36] modeling suite to reflect the mechanism of tetramer formation. Figure 4.1 presents the comparison of the kinetic experimental data to the computational model of tetramer formation. In the figure, the blue line represents the monomer ratio as a function of time, the green line represents the dimer ratio as a function of time, the red line represents the ratio of tetramer as a function of time, and the red x's represent the FRET data points. The computational model predicts the intermediate dimer prior to the tetramer, but also shows that with the concentration evaluated ($1\mu M$ CLZ), the tetrameric state is significantly more favorable than the dimer, and the dimer is readily transitioned into tetramer. In the future, we could try different temperatures to see if we can observe experimentally the two-step process predicted by the simulations.

The computational extrapolation of the K_D was used to produce a plot of percent of tetramers in solution (at equilibrium) as a function of logarithmic CLZ concentration. This figure is compared to the equilibrium FRET ratio data and is presented in Figure 4.2.

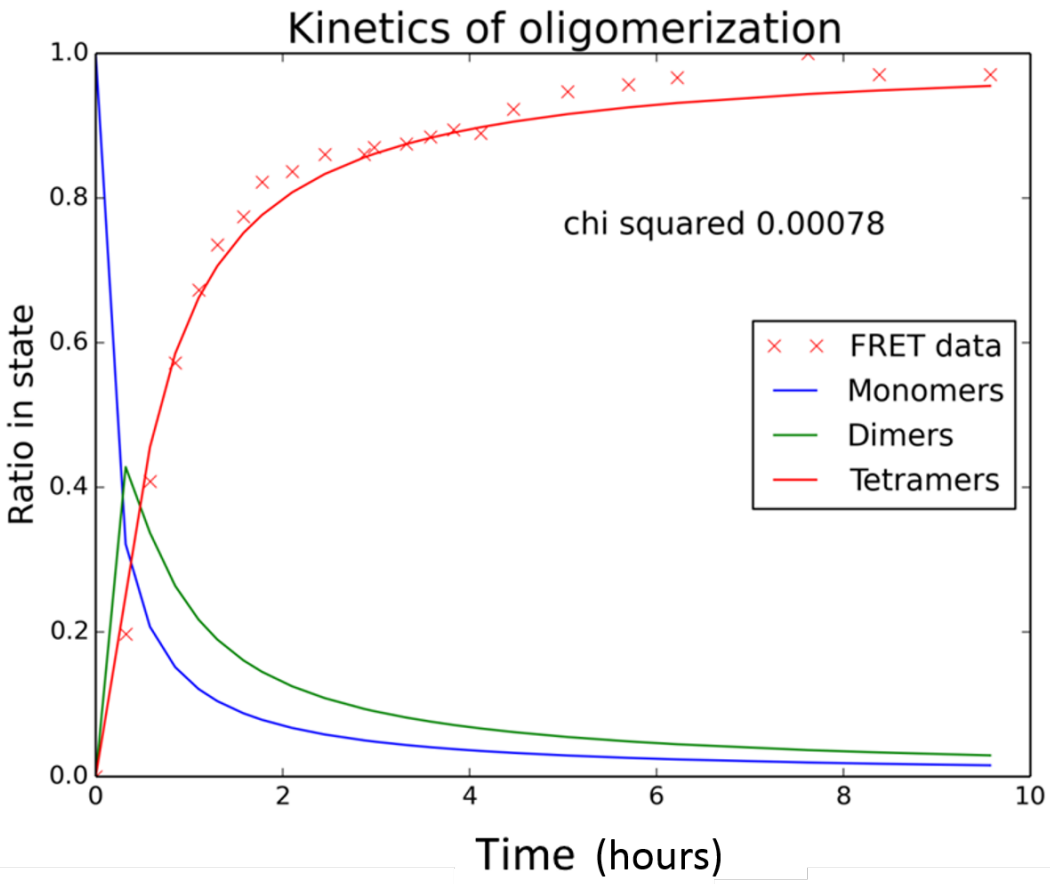

FIGURE 4.1: Computational model of tetramer formation parametrized with kinetic FRET data.

The need for experimental comparison and parameterization of computational work is illustrated by comparing our experimentally found ΔG° to that found by a purely computational study conducted by Lisse *et al.* (2014) [37]. In their study, they determined the ΔG° of hnRNP C CLZ domain tetramerization to be $-112.07\ kcal\ mol^{-1}$ ($-460.91\ kJ\ mol^{-1}$), an extremely large difference for a process that does not involve the formation or breaking of covalent bonds. For comparison, the indexed ΔG°_{f} for liquid water (H_2O) at 1 atm and $25^{\circ}C$ is reported as $-237.2\ kJ\ mol^{-1}$ [38]. The corresponding K_D can be calculated as $5.63 \times 10^{-81} M$, a value that is unobserved in biological systems, and

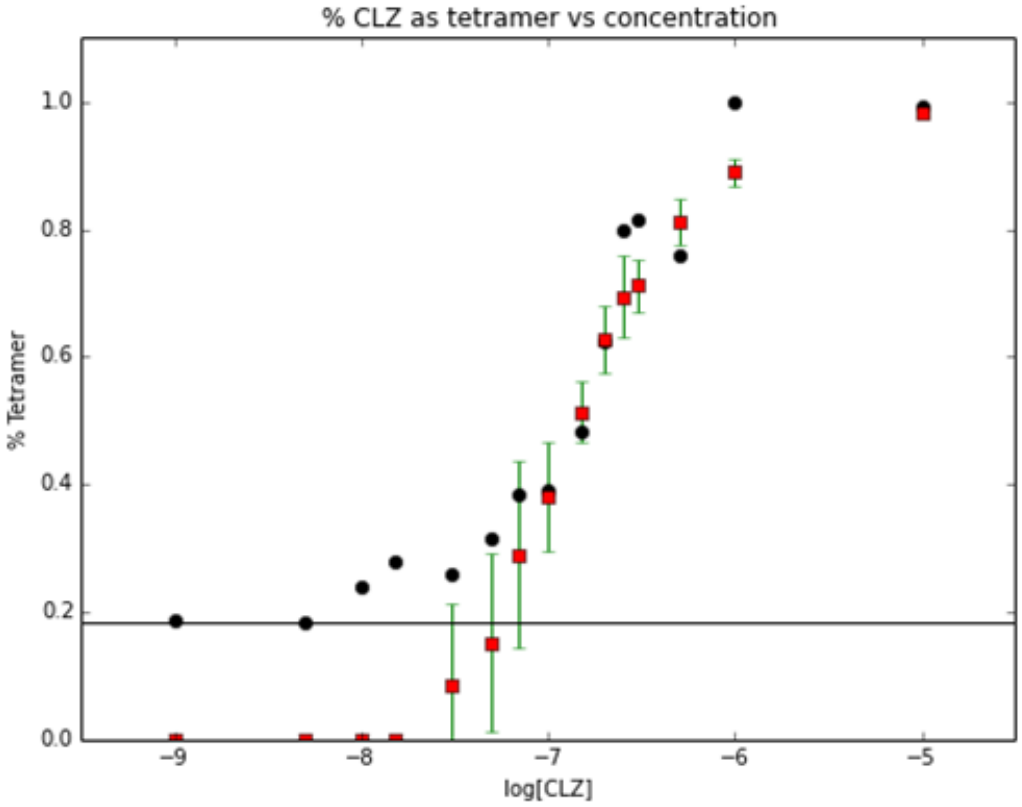

FIGURE 4.2: Computational model of tetramer formation parametrized with equilibrium FRET data. Red squares represent computational data while black circles represent experimental data.

would indicate an essentially irreversible reaction. For reference, many biological associations of macromolecules similar to CLZ have ΔG° values in the range of $-15\ kJ\ mol^{-1}$ to $-45\ kJ\ mol^{-1}$ [16]. Our experimentally determined ΔG° was $-9.0751 \pm 0.0002\ kcal\ mol^{-1}$ $(-37.9712\ \pm 0.0008\ kJ\ mol^{-1})$.

4.2 FRET Discussion

The low CLZ concentration samples had interesting excitation curves (3.2, and tests were done to ensure we were not experimenting below the limit of quantification with the PMTs. This was accomplished by opening up the slit widths of both the excitation and emission slits and by increasing the integration time per wavelength. The increase in slit widths allows higher intensity light to reach the sample and detector, and the increase in integration time allows a better representative sample of each wavelength to be collected for analysis. As a consequence of these measures, the intensity increased for all areas of the spectra, however, the shape of the curve was the same as before. This result infers that at 3 nM we have not yet reached a concentration where the shape of the curve cannot be trusted as representative of what is occurring within the sample.

Previous work in the Whitson research group used fluorescence emission scans to monitor the occurrence of FRET. However, there are several problems associated with using FRET emission scans. The first and most severe problem is dealing with acceptor emission from the incident light [30]. This problem is due to slight overlap between both fluorophore's excitation spectra (Figure 1.4); when trying to excite the donor there is inherent excitation of the acceptor that is not due to FRET. The most common technique to combat this problem is to measure the intensity of acceptor emission without the presence of the donor molecule in solution and subtract the values from the scan with the donor present [24]. A second problem involves the quantum yield of the two fluorophores. For the same concentration, a fluorophore with a high quantum yield will emit more photons than a fluorophore with a lower quantum yield. The acceptor

fluorophore in FRET faces an inherent disadvantage compared to the donor in that the intensity it received is already a fraction of the incident intensity. The acceptor emission intensity can be further reduced compared to the donor emission if the quantum yield of the acceptor is naturally lower than that of the donor.

For this experiment, the quantum yield of our acceptor, tetramethylrhodamine (TMR), is significantly lower than the quantum yield of the donor fluorescein. While tetramethylrhodamine in solution has a high quantum yield, when complexed to proteins its excitation spectra shifts. Figure 4.3 presents the absorbance scans obtained when tetramethylrhodamine is by itself (TMR), when tetramethylrhodamine is bound to goat anti-mouse IgG antibody (TMR-GAM), and when the TMR-GAM complex is treated with 4.8 M guanidine hydrochloride and the protein complex is denatured. The plot infers that tetramethylrhodamine in a covalently-conjugated state emits less intensity when excited at 550 nm than when in solution alone. This is one possible explanation for why the emission peak in FRET is at such a low intensity compared to the fluorescein peak. In instances of FRET, the emission shoulder of fluorescein covered the emission peak of tetramethylrhodamine. To address this problem we utilized excitation scans to probe the occurrence of FRET. When using excitation scans, the fluorimeter detector is focused on an acceptor emission wavelength, and the incident light is scanned through the range of wavelengths that excite the donor. Without FRET, the excitation of the donor fluorophore doesn't induce emission from the acceptor. With FRET, however, the excitation of the donor increases the acceptor emission, and the increase in photons detected when exciting the donor can be linked to FRET occurrence. The work shown

here reveals that using excitation scans can be just as informative as emission scans when performing FRET experiments with a low quantum yield fluorophore.

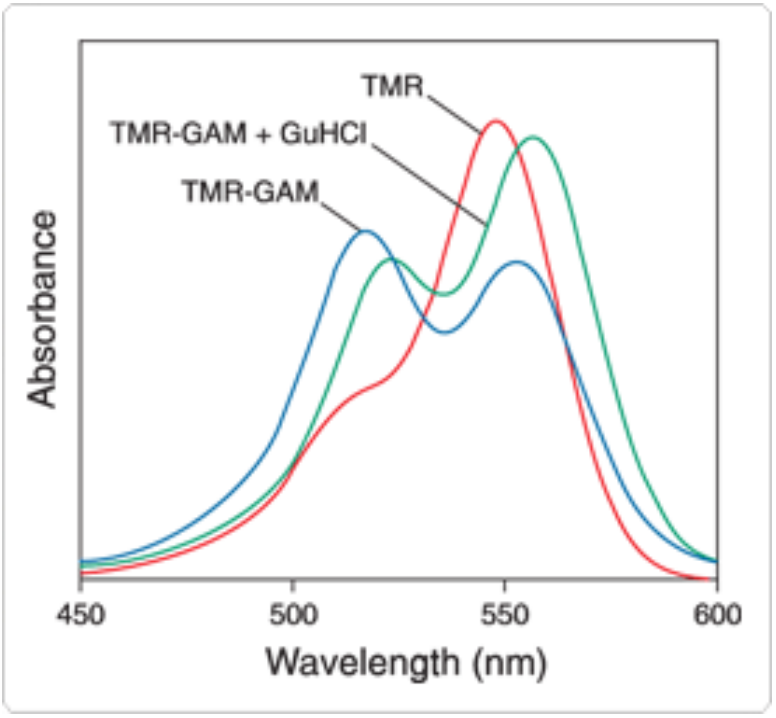

FIGURE 4.3: Effect of protein conjugation on TMR[39].

4.3 Future Direction

Future studies provide the opportunity to further explore the thermodynamic stability of the hnRNP C oligomerization domain. Further equilibrium experiments performed at various temperatures would provide the opportunity to define the standard-state enthalpy (ΔH°) and entropy (ΔS°) for the tetramerization process as well as experimentally determine ΔG°. Further kinetic experiments will allow for better parameterization with the computational model.

4.4 Conclusion

The hnRNP C performs a nonredundant role during the process of pre-mRNA processing. The proper association of the hnRNP C complex relies on the formation of the CLZ coiled coil tetramer. Coiled coil domains represent some of the most stable protein structures, with the CLZ tetramer resisting denaturation up to $89°C$ [2]. This thesis explored computational and experimental methods for investigating the thermodynamic properties of this complex. FRET was the experimental tool of choice and allowed for determination of when fluorophore-labeled CLZ helices formed tetramers. Equilibrium experiments were done to determine the dissociation constant of the CLZ tetramer (K_D). Kinetic experiments were conducted to measure the rate of tetramer formation for several concentrations. Mathematical models of the process were made in conjunction with collaborators at Vanderbilt University, and experimental data was used to parameterize the models. The tetramerization mechanism predicted by the computational model fit the experimental data very well, and provided insight where traditional data fitting methods failed. Future work will further investigate the thermodynamic properties of the hnRNP C oligomerization domain using novel computational and experimental approaches.

Appendix A

Equilibrium Dependence on Subunit Concentration

This appendix discusses the role that subunit concentration plays in the formation of complex. The following derivation elucidates in a clear manner why CLZ tetramers are not expected to form when total CLZ concentration is below the equilibrium constant (K_D). The process of this derivation is very similar to that of Van Holde (2006), however it uses the equilibrium dissociation constant rather than the equilibrium association constant[16]. This derivation examines a simple model of two monomers associating to form a dimer,

$$\text{Monomer} + \text{Monomer} \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} \text{Dimer}$$

When this reaction is at equilibrium,

$$k_{on}[M]^2 = k_{off}[D] \tag{A.1}$$

which can be rearranged for the equilibrium dissociation constant:

$$K_D = \frac{k_{off}}{k_{on}} = \frac{[M]^2}{[D]} \tag{A.2}$$

Now we allow S_0 to represent the total molar concentration of subunits in the system, which can be represented by Equation A.3. In this expression, the factor of 2 appears due to each dimer containing 2 moles of the subunit $[M]$. Equation A.3 can be rearranged for the dimer concentration ($[D]$).

$$S_0 = [M] + 2[D] \tag{A.3}$$

$$[D] = \frac{S_0 - [M]}{2} \tag{A.4}$$

Equation A.4 can be substituted into Equation A.2, which can then be solved using a quadratic equation to yield an expression for [M]. This resulting expression can be divided by S_0 to yield Equation A.5, a relation of monomer concentration as a function of subunit concentration.

$$\frac{[M]}{S_0} = \frac{-K_D + K_D\sqrt{1 + \frac{8S_0}{K_D}}}{4S_0} \tag{A.5}$$

When the limits of this expression are evaluated, and a binomial expansion is applied to the term inside the square root, it is observed that as $S_0 \to 0$, $\frac{[M]}{S_0}$ goes to 1.

$$\lim_{S_0 \to 0} \frac{[M]}{S_0} = \frac{K_D}{4S_0}(-1 + \sqrt{1 + \frac{8S_0}{K_D}})$$

$$= \frac{K_D}{4S_0}(-1 + 1 + \frac{8S_0}{2K_D})$$

$$= \frac{K_D}{4S_0}(\frac{4S_0}{K_D}) = 1$$

Evaluating the other limit shows that as $S_0 \to \infty$, $\frac{[M]}{S_0}$ goes to 0. As the concentration of total subunits in the sample increases, the ratio of monomers in the sample decreases and all the subunits form the dimer. If the concentration of total subunits in the sample decreases, the sample consists of mainly monomers.

Bibliography

[1] P. J. Russell. *iGenetics; A Molecular Approach.* Pearson Education Inc., 2nd edition, 2006.

[2] S. R. Whitson, W. M. LeStourgeon, and A. M. Krezel. Solution structure of the symmetric coiled coil tetramer formed by the oligomerization domain of hnrnp c: Implications for biological function. *Journal of Molecular Biology*, 350(2):319–337, 2005.

[3] S. Pinol-Roma, Y. D. Choi, M. J. Matunis, and G. Dreyfuss. Immunopurification of heterogeneous nuclear ribonucleoprotein particles reveals an assortment of rna-binding proteins. *Genes Dev*, 2:215–227, 1988.

[4] A. M. Krecic and M.S. Swanson. hnrnp complexes: composition, structure, and function. *Curr. Opin. Cell Bio*, 11:363–371, 1999.

[5] S. Nakielny and G. Dreyfuss. The hnrnp c proteins contain a nuclear retention sequence that can override nuclear export signals. *J. Cell Bio*, 134(6):1365–1373, 1996.

[6] A. McCloskey, I. Taniguchi, K. Shinmyozu, and M. Ohno. hnrnp c tetramer measures rna length to classify rna polymerase ii transcripts for export. *Science*, 335:1643–46, March 2012.

[7] C. G. Burd, M. S. Swanson, M. Gorlach, and G. Drefuss. Primary structures of the heterogeneous nuclear ribonucleoprotein a2, b1, and c2 proteins: a diversity of rna binding proteins is generated by small peptide inserts. *Proc NAtl Acad Sci U S A*, 86:9788–9792, 1989.

[8] M. Buvoli, F. Cobianchi, M. G. Bestagno, A. Mangiarotti, M. T Bassi, G. Biamonti, and S. Riva. Alternative splicing in the human gene for the core protein a1 generates another hnrnp protein. *Embo J*, 9:1229–1235, 1990.

[9] M. S. Swanson, T. Y. Nakagawa, K. LeVan, and G. Dreyfuss. Primary structure of human nuclear ribonucleoprotein particle c proteins: conservation of sequence and domain structures in heterogeneous nuclear rna, mrna, and pre-rrna-binding proteins. *Mol Cell Bio*, 7:1731–1739, 1987.

[10] S. F. Barnett, D. L. Friedman, and W. M. LeStourgeon. The c protins of hela 40s nuclear ribonucleoprotein particles exist as anisotropic tetramers of (c1)3 c2. *Mol Cell Biol*, 9:492–498, 1989.

[11] M. Huang, J. E. Rech, S. J. Northington, P. F. Flicker, A. Mayeda, A. R. Krainer, and W. M. LeStourgeon. The c-protein tetramer binds 230 to 240 nucleotides of pre-mrna and nucleates the assembly of 40s heterogeneous nuclear ribonucleoprotein particles. *Mol Cell Biol*, 14:518–533, 1994.

[12] J. E. Rech, M. H. Huang, W. M LeStourgeon, and P. F. Flicker. An ultrastructural characterization of in vitro-assembled hnrnp c protein-rna complexes. *J Struct Biol*, 114:84–92, 1995.

[13] L. Pauling, R. B. Corey, and H. R. Branson. The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. *Proc Natl Acad Sci U S A*, 37:205–211, 1951.

[14] C. Branden and J. Tooze. *Introduction to Protein Structure*. Garland Science, New York, 2 edition, 1999.

[15] A. L. Morris, M. W. MacArthur E. G. Hutchinson, and J. M. Thornton. Stereochemical quality of protein structure coordinates. *Proteins*, 12:345–364, 1992.

[16] K. E. Van Holde, C. Johnson, and P. S. Ho. *Principles of Physical Biochemistry*. Prentice Hall, New Jersey, 2 edition, 2006.

[17] L. Shahied, E. H. Braswell, W. M. LeStourgeon, and A. M. Krezel. An antiparallel four-helix bundle orients the high-affinity rna binding sites in hnrnp c: a mechanism for rna chaperonin activity. *J Mol Biol*, 305:817–828, 2001.

[18] J. G. McAfee, L. Shahied-Milam, S. R. Soltaninassab, and W. M. LeStourgeon. A major determinant of hnrnp c protein binding to rna is a novel bzip-like rna binding domain. *RNA*, 2:1139–52, 1996.

[19] A. M. Friedman, T. O Fischmann, and T. A. Steitz. Crystal structure of lac repressor core tetramer and its implications for dna looping. *Science*, 268:1721–1727, 1995.

[20] G. M. Clore, J. G. Omichinski, K. Sakaguchi, N. Zambrano, H. Sakamoto, E. Appella, and A. M. Groenenborn. High-resolution structure of the oligomerization domain of p53 by multidimensional nmr. *Science*, 265:386–391, 1994.

[21] R. Eisberg and R. Resnick. *Quantum Physics of Atoms, Molecules, Solids, Nuclei, and Particles*. John Wiley and Sons, New York, 2 edition, 1985.

[22] J. D. Cutnell and K. W. Johnson. *Physics*. Wiley, New York, 9 edition, 2012.

[23] D. Skoog, F.J. Holler, and S.R. Crouch. *Instrumental Analysis*. CENGAGE Learning, New York, 2007.

[24] J. R. Lakowicz. *Principles of Fluorescence Spectroscopy*. Plenum Publishers, New York, 2nd edition, 1999.

[25] I. Johnson and M. T. Z. Spence. *Molecular Probes Handbook, A Guide to Fluorescent Probes and Labeling Technologies*. Life Technologies, 9 edition, 2010.

[26] J. C. Pino, B. Kim, S. R. Whitson, and K. B. Whitson. Measurement of coiled-coil stability by fluorescence resonance energy transfer. *Amer. Phys. Soc.*, 57(36), 2012.

[27] P. Atkins and J. D. Paula. *Physical Chemistry*. W. H. Freeman and Company, New York, 9th edition, 2010.

[28] I. Z. Steinberg. Long-range nonradiative transfer of electronic excitation energy in proteins and polypeptides. *Annu. Rev. Biochem.*, 40:83–114, 1971.

[29] L. Stryer. Fluorescence energy transfer as a spectroscopic ruler. *Annu. Rev. Biochem.*, 47:819–846, 1978.

[30] J. R. Lakowicz, I. Gryczynski, W. Wiczk, G. Laczko, F. C. Prendegast, and M. L. Johnson. Conformational distributions of melittin in water/methanol mixtures from frequency-domain measurements of nonradiative energy transfer. *Biophys. Chem.*, 36:99–115, 1990.

[31] D. A. Johnson, V. L. Leathers, A. M. Martinez, D. A. Walsh, and W. H. Fletcher. Biomedical example: Use of fret to measure subunit associations of the regulation (r) and catalytic (c) subunits of a protein kinase. *Biochemistry*, 32:6402–6410, 1993.

[32] S. S. Ghosh, P. S. Eis, K. Blumeyer, K. Fearon, and D. P. Millar. Real time kinetics of restriction endonuclease cleavage monitored by fluorescence resonance energy transfer. *Nucleic Acids Res.*, 22:3155–3159, 1994.

[33] H. Motulsky. *The GraphPad Guide to Analyzing Radioligand Binding Data*. GraphPad Software, 1996.

[34] S. Pronk et al. Gromacs 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 2013.

[35] S. Kumar et al. The weighted historgram analysis method for free-energy calculations on biomolecules. i. the method. *Journal of computational Chemistry*, 13(8), 1992.

[36] J. A. Sekar and J.R. Faeder. Rule-based modeling of signal transduction: a primer. *Methods Mol. Biol.*, 880:139–218, 2012.

[37] T.S. Lisse, K. Vadivel, S. P Bajaj, R. F. Chun, M. Hewison, and J. S. Adams. The heterodimeric structure of heterogeneous nuclear ribonucleoprotein c1/c2 dictates

1,25-dihydroxyvitamin d-directed transcriptional events in osteoblasts. *Bone Res.*, 2, 2014.

[38] J. Burdge and J. Overby. *Chemistry: Atoms First*. McGraw Hill, New York, NY, 2015.

[39] D. L. Meadows, J. S. Shafer, and J. S. Schultz. Determining the extent of labeling for tetramethylrhodamine protein conjugates. *J Immunol Methods*, 143(2):263–272, 1991.